On Using Gait to Enhance Frontal Face Extraction

Sung-Uk Jung and Mark S. Nixon

***Abstract*—Visual surveillance finds increasing deployment for monitoring urban environments. Operators need to be able to determine identity from surveillance images and often use face recognition for this purpose. In surveillance environments, it is necessary to handle pose variation of the human head, low frame rate, and low resolution input images. We describe the first use of gait to enable face acquisition and recognition, by analysis of 3-D head motion and gait trajectory, with super-resolution analysis. We use region- and distance-based refinement of head pose estimation. We develop a direct mapping to relate the 2-D image with a 3-D model. In gait trajectory analysis, we model the looming effect so as to obtain the correct face region. Based on head position and the gait trajectory, we can reconstruct high-quality frontal face images which are demonstrated to be suitable for face recognition. The contributions of this research include the construction of a 3-D model for pose estimation from planar imagery and the first use of gait information to enhance the face extraction process allowing for deployment in surveillance scenarios.**

***Index Terms*—Biometrics, front face extraction, gait trajectory, head pose estimation, looming effect.**

I. Introduction

As Closed-Circuit Television (CCTV) becomes more widespread, it can be used for authentication within visual surveillance; however, developments have yet to meet application requirements. For example, the police still use manual search tactics to find criminals in CCTV images; airports depend on documents, although biometrics is becoming more prevalent. Automatic face recognition potentially has a major role to play in surveillance, information security, and access control. To date, automatic face recognition has mainly used 2-D front-face pattern and texture information. Even when these are confined to indoor usage, the resulting recognition rate is not perfect, although performance continues to improve. The reasons for this improvement include the use of very high resolution frontal face images in constrained environments [1] and more sophisticated recognition approaches. In contrast, in visual surveillance environments a camera is in an elevated position to achieve the widest field of view. Many surveillance cameras have low resolution and low frame rate, constraining adoption of existing face tracking/recognition algorithms.

Manuscript received March 07, 2012; revised August 12, 2012; accepted September 02, 2012. Date of publication September 12, 2012; date of current version November 15, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ajay Kumar.

S.-U. Jung is with the Cyber Security-Convergence Research Division, ETRI, Daejeon 305-500, Korea (e-mail: brcastle@etri.re.kr).

M. S. Nixon is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: msn@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2012.2218598

There are many approaches applied to obtain High Resolution (HR) face images. Medioni *et al.* [2] detected a person and located their face using a fixed ultrahigh-resolution camera at a distance (3, 6, 9 m). Using the face region, they framed and reconstructed a 3-D face model and recognized a face within it. Wheeler *et al.* [3] initialized Low Resolution (LR) face images using an Active Appearance Model (AAM) [25] and Bilateral Total Variation (BTV) [42] to generate HR images. This process was tested with outdoor videos using a PTZ camera and a commercial face recognition system (FaceIt—Identix). Mortazavian *et al.* [4] described an example-based Bayesian method for 3-D-assisted pose-independent face texture super-resolution. They used a 3-D Morphable Model (3DMM) [14] to map facial texture from a 2-D face image. Jia and Gong [5] proposed learning-based face Super Resolution (SR) techniques to generate a HR image of a single facial modality such as a fixed expression, pose and illumination from given LR images. There are other studies which learned the relation between HR and LR and utilized the relationship to recognize the LR face images from visual surveillance data [6]–[8].

Also, 3-D model-based approaches have been developed to obtain an accurate face model [9], [10]. Park and Jain [11] initialized face images using an AAM and synthesized a 3-D frontal face from 2-D low resolution images. They demonstrated performance using commercial recognition software (FaceVACS—Cognitec) and the Face In Action (FIA) database. This system was a single camera-based system requiring initialization. Duhn *et al.* [12] generated a 3-D face model using three 2-D images and tracked a face using an AAM. In the tracking stage, a generic model was adapted to the different views of the faces. However this method still needed automatic initialization for tracking. Given 2.5-D face laser scan images, Lu *et al.* [13] constructed a 3-D head model by fitting the laser scan data with the pretrained face texture and shape. They used Iterated Closest Points (ICP)-based surface matching to construct the 3-D model; however, this approach suffered from the computational cost of 3-D data acquisition and processing. The most representative method of 3-D face modeling is a 3DMM [14] wherein a 3-D face model was constructed by using a form of 3-D Principal Components Analysis (PCA). The coefficients of texture and shape are used for face recognition. However, both training and test images need to be manually labeled and the cost of 3-D data acquisition and processing is high. Liao *et al.* [15] built a 3DMM based on a single image to recognize a person. They generated synthetic 2-D images in the training stage and recognized a face using a Support Vector Machine (SVM) with imposter rejection by local linear embedding. In the CHIL project [16] the goal of the project was to acquire information about the (smart) room by video surveillance of the people in it and their interaction. The tasks of this project included person tracking, person identification and head pose

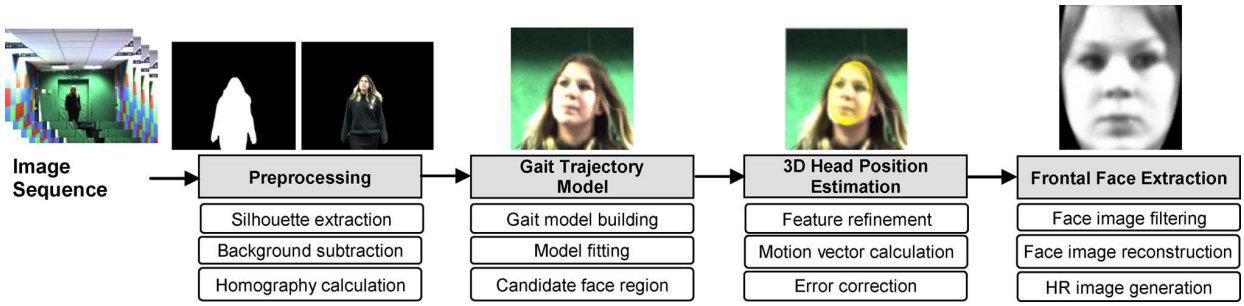


Fig. 1. Front-face acquisition process.

estimation, and the recognition rate of individuals was from 66.2% to 75.9% using low resolution face images.

There are other approaches in visual surveillance using multimodal biometrics—face and gait. Human gait analysis has been used for recognition and other applications such as age estimation and authentication [17]–[19]. Since the most distinguishable feature in these environments is the human gait, earlier approaches tried to combine these two modalities to improve recognition. Kale *et al.* [20] discussed the fusion of face and gait cues for a single camera, based on sequential importance sampling. Gait recognition was used to reduce the set of candidates for face recognition. The results of fusion experiments were demonstrated on the NIST database which had outdoor gait and face data of 30 subjects [21] and suggested that a multimodal biometric could achieve a higher recognition rate than a single-modal biometric. Shakhnarovich *et al.* [22] developed a view-normalization approach to multiview face and gait recognition. In this approach, the Image-Based Visual Hull (IBVH) was computed from a set of monocular views and used to render virtual views. Zhou *et al.* [23] combined cues of face profile and gait silhouette from a single camera video sequence. First, they reconstructed a high resolution face profile image and used a Gait Energy Image (GEI) to characterize human walking properties.

This paper is an extended version of our previous research [30], [31]. We now analyze the 3-D face model in more detail and the extraction of the gait characteristics, together with performance comparison with other techniques. In this paper, we focus on the specific but also one of the most general environments in visual surveillance where the subject is viewed walking towards a camera. We suppose that the image sequences are recorded at low frame rate and low resolution. To overcome these constraints, a 3-D ellipsoidal head pose model is built, and estimate 3-D head pose using nonlinear optimization. However, the head pose model can fail to track the head when there is large head pose variation and changes in illumination. Thus, we use additional information—the gait trajectory. The gait trajectory can be obtained from the movement of the corresponding points between frames. Analyzing the gait trajectory allows the system to detect the approximate face regions. In these face regions the detection rate can be higher. Accordingly, this is the first work which uses gait to mediate face acquisition and face extraction enhancement.

Fig. 1 describes the whole process; the remainder of this paper is organized as follows: Section II describes the head pose estimation method using a 3-D ellipsoidal model and nonlinear optimization. Modeling the gait trajectory and optimization are described in Section III. Section IV shows the process of the frontal face extraction and the experimental results. Finally, Section V discusses the conclusions and future directions of this study.

II. 3-D Head Pose Estimation

There are many approaches to estimate 3-D head pose. Some studies use a real 3-D head model [14], [25] and others use an approximated head model [27]–[29]. Basically, estimating the head pose requires calculation of the translation and rotation in 3-D space. If the 3-D corresponding points exist and there is a small movement between images, the motion vector can be calculated by using the twist representation [32] or optical flow. Another way is to use Epipolar geometry [33] to obtain the essential matrix from which the rotation and translation information can be extracted. However, the first calculates the motion vector based on the angle approximation in 3-D space and the second can work when the texture of image is clear so the optical flow can be calculated. The third method could be unstable if the corresponding points used to calculate the fundamental matrix lie on the same plane. Accordingly, none of the above methods is suitable for our application, motivating us to apply a new way to estimate the head pose. First of all, the model-based feature extraction method is described. Then, a motion vector is estimated based on an objective function which describes the relationship between reconstructed 3-D points and 2-D corresponding points by using nonlinear optimization. Finally, the error is corrected by the analysis of the 2-D representation of 3-D model.

A. Feature Refinement Aided by the Model

A 3-D ellipsoidal representation [28] is used to represent a face. Unlike a cylinder model, the ellipsoidal model considers the distinguishable part of the face, excluding the background. A point on a 3-D object $\mathbf{P}_o$ can be represented by $[X_o Y_o Z_o]$where $X_o = r_x sin\alpha\ sin\beta$, $Y_o = r_y cos\alpha$, and $Z_o = r_z sin\alpha\ cos\beta$, r is the radius along each axis. Each angular resolution is one degree, so that the total number of model components is 360×360. The Scale Invariant Feature Transform (SIFT) [35] is deployed to find corresponding points between two images. We calculate a 3-D rotation and translation matrix from the SIFT points matched between adjacent images.

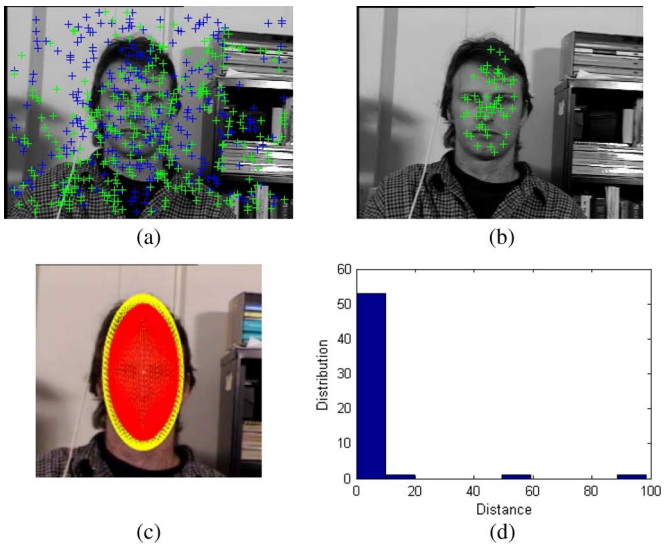


Fig. 2. Feature filtering. (a) Features before filtering; (b) features after filtering; (c) confidence region; (d) distance histogram.

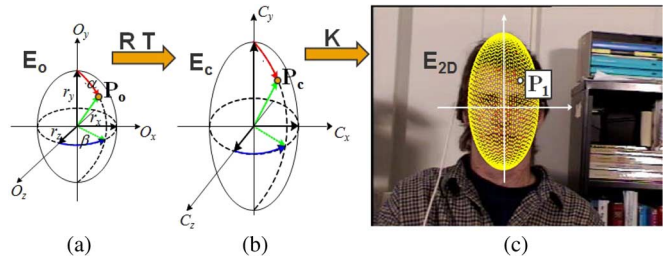


Fig. 3. Direct mapping from 2-D image to 3-D model. (a) Object coordinate; (b) camera coordinate; (c) image coordinate.

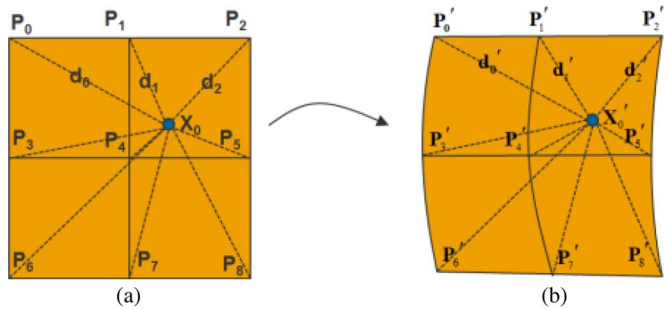


Fig. 4. Extracted SIFT point and around points in (a) 2-D and (b) 3-D.

Once the initialization process is complete, which requires manual specification of the rotation and translation of the model in the first frame, the invalid features should be removed. Fig. 2(a) shows the extracted SIFT features. The green crosses depict the SIFT points which are matched to the following frame. The corresponding points selected by the SIFT descriptor are refined by a region-based and a distance-based measure because misalignment could occur since matching with only the SIFT descriptor considers the pattern around the extracted SIFT point.

Therefore, first, the SIFT matching points only within $\pm 50^\circ$ from the center point ($\alpha = 90^\circ, \beta = 0^\circ$ in the ellipsoidal model shown in Fig. 3(a)) are considered. Fig. 2(c) displays the region. The yellow region is the fitted model, the red region shows the region within $\pm 50^\circ$ from the center point, and the white cross in the model is the center point. As shown in the figure, the red region can cover the facial components (eyes, nose and mouth) which hold rich information on head pose. The second is to filter the invalid matching points based on the distances between each pair of potential corresponding points. The distances can show the distribution of movements. In histogram of the distances, the misaligned points can be detected and removed. In an alternative representation, the distribution of histogram can be expressed by the Gaussian distribution ($X \sim N(\mu, \sigma^2)$). The SIFT points are taken between $\mu - 2\sigma$ and $\mu + 2\sigma$, covering 68.2% of the Gaussian distribution. Fig. 2(d) shows the distance distribution from each pair of matched SIFT points. The final valid features are shown in Fig. 2(b).

After filtering, the 3-D positions of SIFT points in the object coordinates are reconstructed by direct mapping. As shown in Fig. 3, the 3-D point on the ellipsoidal model is mapped to the 2-D point in the image via rotation, translation, and projection. In Fig. 3, a model point $\mathbf{P}_\mathrm{o}$ is mapped to the point $\mathbf{P}_\mathrm{c}$ by changing from the object coordinates to the camera coordinates. Then, it is mapped to the point $\mathbf{P}_1$ in the image plane by the camera matrix. Conversely, if the point $\mathbf{P}_1$ is known, the point $\mathbf{P}_\mathrm{o}$ (in the object coordinates) can be found directly. Therefore, assuming that the point $\mathbf{P}_1$ is an extracted SIFT point, the 3-D position in the object coordinate can be calculated.

Practically, the ellipsoidal model $\mathbf{E}_\mathrm{o}$ and $\mathbf{E}_\mathrm{c}$ have 360×360 (spherical) points with one degree resolution. For the projected ellipsoidal model $\mathbf{E}_{2\mathrm{D}}$, only 180×360 points are valid due to invisible parts. The extracted SIFT point could be one of point in the model $\mathbf{E}_{2\mathrm{D}}$ or between points. Therefore, to reconstruct an exact 3-D position in the model $\mathbf{E}_\mathrm{o}$ we approximately linearize the position using the distance ratio between the extracted SIFT point and around nine points on the model. First, the closest point of the model to the SIFT point is found, and the nine points around the SIFT point are chosen. Then, the distance ratios between the SIFT point and around the points are calculated. The same procedure is applied to find 3-D position in the object coordinates as the ratio should be consistent. The corresponding points in the object coordinate for all of the 2-D points can be found. Then, using the distance ratio, the approximated 3-D position of 2-D SIFT points can be calculated.

Fig. 4 shows this relationship. $\mathbf{X}_\mathrm{o}$ and $\mathbf{P}_\mathrm{n}$ are extracted SIFT points and the points in the model $\mathbf{E}_{2\mathrm{D}}$. $\mathbf{X}'_0$ and $\mathbf{P}'_\mathbf{n}$ are the corresponding points in the model $\mathbf{E}_\mathrm{o}$. Let the distance between $\mathbf{X}_\mathrm{o}$ and $\mathbf{P}_\mathrm{n}$ be d_n; thus

$$r_k = (1/d_k) \Big/ \sum_{k=1}^{n} 1/d_k \quad \text{then} \quad \sum_{k=1}^{n} r_k = 1 \tag{1}$$

The 3-D position, $\mathbf{X}'_0$ can be calculated by

$$\mathbf{X}'_0 = \sum_{k=1}^{n} r_k \cdot \mathbf{P}'_k \tag{2}$$

From the above relationship the 3-D position is reconstructed and this 3-D position will be used to calculate the motion vector in the next Section.

B. Motion Vector Estimation Using Nonlinear Optimization

The Levenberg-Marquadt algorithm is a nonlinear optimization algorithm which can provide the numerical solution to minimize an objective function. We define the objective function to extract the motion information as

$$\arg\min_{\mathbf{R}\ \mathbf{T}} \sum_{k=1}^{n} \left\| \mathbf{p}'_{2d,k} - \mathbf{K}(\mathbf{R}\mathbf{P}_{3d,k} + \mathbf{T}) \right\| \tag{3}$$

where $\mathbf{p}'_{2d,k}$ is a SIFT point in the next frame, and $\mathbf{P}_{3d,k}$ is a reconstructed 3-D point from the corresponding point in the current frame. n is the total number of the pairs. The rotation matrix $(\mathbf{R})$ contains $\Delta\theta_x, \Delta\theta_y, \Delta\theta_z$ and the translation matrix $(\mathbf{T})$ contains $\Delta x, \Delta y, \Delta z$. In (3), $\mathbf{P}_{3d,k}$ is a position of point in the object-oriented coordinate and $\mathbf{p}'_{2d,k}$ is a position of SIFT point in the 2-D image space. Also, we assume the camera matrix $(\mathbf{K})$ is given. Basically, (3) finds the rotation and translation matrices which minimize the distance between the matched points in 2-D space. The term $\mathbf{K}(\mathbf{R}\mathbf{P}_{3d,k} + \mathbf{T})$ in (3) converts the point from the coordinate 3-D space into 2-D image space. The motion information is extracted to adapt this objective function using the Levenberg-Marquadt algorithm. In this way, the rotation and translation matrix can be updated using the previous motion information for each image frame.

C. Error Correction

In the previous Section, the motion vector is calculated. However, since the model we used is approximate model (assuming that the head shape is an ellipse) an accumulated tracking error and initialization error could occur according the frame number. Therefore, an error correction module is necessary. If assuming that variation between the potential motion vector calculated previously and the real motion vector is quite small, Optical flow [34] can be used to correct the translation vector of the motion vector. Under small motion variation and no illumination change the velocity relationship between 2-D and 3-D motion can be described as the following equation [34],

$$\frac{1}{Z}\left[fI_x fI_y - (xI_x + yI_y)\right] \mathbf{R}\left[\mathbf{I} - [\mathbf{P}_{\mathrm{o}}]_{\times}\right] \begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{r} \end{bmatrix} = -I_t \tag{4}$$

where I_x, I_y and I_t are image intensity gradients with respect to x, y, and t respectively. $\mathbf{P}_{\mathrm{o}}$ is a point in the object coordinate. $\mathbf{R}$, $\Delta\mathbf{t}$, and $\Delta\mathbf{r}$ are rotation matrix in 3-D and instantaneous translation and rotation vector, respectively. $[\mathbf{I} - [\mathbf{P}_{\mathrm{o}}]_{\times}]$ is a matrix formed by concatenating $\mathbf{I}$ and $-[\mathbf{P}_{\mathrm{o}}]_{\times}$. $[]_{\times}$ denotes a skew-symmetric matrix. $\mathbf{I}$ is an identity matrix and f is a focal length. A linear equation can be made by adapting (4) to all visible pixels. Then, a least squares solution is used to calculate the translation and rotation vector $(\Delta\mathbf{t}, \Delta\mathbf{r})$. To deploy (4) the rotation matrix $\mathbf{R}, \mathbf{T}$ are changed into $\mathbf{R_p}, \mathbf{T_p}$ which are the calculated rotation and translation matrices described in previous Section. Thus, the instantaneous rotation vector can be small.

We also correct the rotation of the motion vector. Our model is the 3-D ellipsoidal model whose resolution is one degree. Once the model is fitted to the image plane a 2-D texture map can be obtained by unfolding the 3-D ellipsoidal model per each frame. Fig. 5 shows the unfolded image in the range of $0° \leq \alpha \leq 180°$, $-90° \leq \beta \leq 90°$. This is only dependent on angles so that the image size is 180 × 180 pixels regardless of the actual head size. Ideally, if the frame rate is high enough, the adjacent unfolded images could have the same texture. Using these texture maps, a 2-D similarity transformation matrix between the adjacent images can be calculated. The procedure is similar to that used to calculate the 3-D motion vector in the sense to use the corresponding SIFT points and nonlinear optimization. First, the corresponding points is extracted using SIFT on the confidence region ($\pm 50°$ from the center point) Then, the parameters of the similarity transform are calculated by using nonlinear optimization. When $\mathbf{x}_{\mathrm{k}}$ and $\mathbf{x}'_{\mathbf{k}}$ are corresponding point between images the similarity transform $(\mathbf{H}_s)$ and the objective function are described as follows,

(a) (b)

Fig. 5. Unfolded images. (a) frame 1; (b) frame 2.

$$\mathbf{x}'_k = \mathbf{H}_s\mathbf{x}_k = \begin{bmatrix} s_{2D}\mathbf{R}_{2D} & \mathbf{T}_{2D} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x}_k \tag{5}$$

$$\arg\min_{\mathbf{R}_{2D},\mathbf{T}_{2D},s_{2D}} \sum_{k=1}^{n} \left\| \mathbf{x}'_k - \mathbf{H}_s\mathbf{x}_k \right\| \tag{6}$$

where s_{2D} is a scaling factor, $\mathbf{R}_{2D}$ is a 2 × 2 rotation matrix, $\mathbf{T}_{2D}$ is a translation 2-vector, and $\mathbf{0}$ is a null 2-vector.

There is a relationship between 2-D motion vector and 3-D motion vector. Let 2-D and 3-D motion vectors be $\mathbf{M}_{2D}, \mathbf{M}_{3D}$

$$\mathbf{M}_{2D} = [s_{2D}, t_{2Dx}, t_{2Dy}, \theta_{2D}] \tag{7}$$

$$\mathbf{M}_{3D} = [t_{3Dx}, t_{3Dy}, t_{3Dz}, \theta_{3Dx}, \theta_{3Dy}, \theta_{3Dz}] \tag{8}$$

The corrected motion vector is

$$\mathbf{M}'_{3D} = [t_{3Dx}, t_{3Dy}, t_{3Dz} \times s_{2D}, \theta_{3Dx} + t_{2Dy}, \theta_{3Dy} + t_{2Dx}, \theta_{3Dz} + \theta_{2D}] \tag{9}$$

Here, the y axis translation in 2-D (t_{2Dy}) is a variation in the direction of the x axis rotation in 3-D space, the x axis translation in 2-D (t_{2Dx}) is a variation in the direction of the y axis rotation in 3-D space. The scaling factor (s_{2D}) is directly matched to the z axis translation.

To verify the performance of face tracking we used the Boston face database [27] which has 45 image sequences plus the ground truth of 3-D head pose. The experimental procedure is as follows. First, corresponding SIFT points are extracted, and then 3-D points in the first frame are calculated by given a motion vector. Then, the potential motion vector is calculated using nonlinear optimization between the 2-D SIFT points and the corresponding 3-D SIFT points. From every motion vector, the unfolded image is generated from the fitted ellipsoidal model. Then, the final motion vector is corrected by the optical flow and the similarity matrix. In Fig. 6, the first row shows the

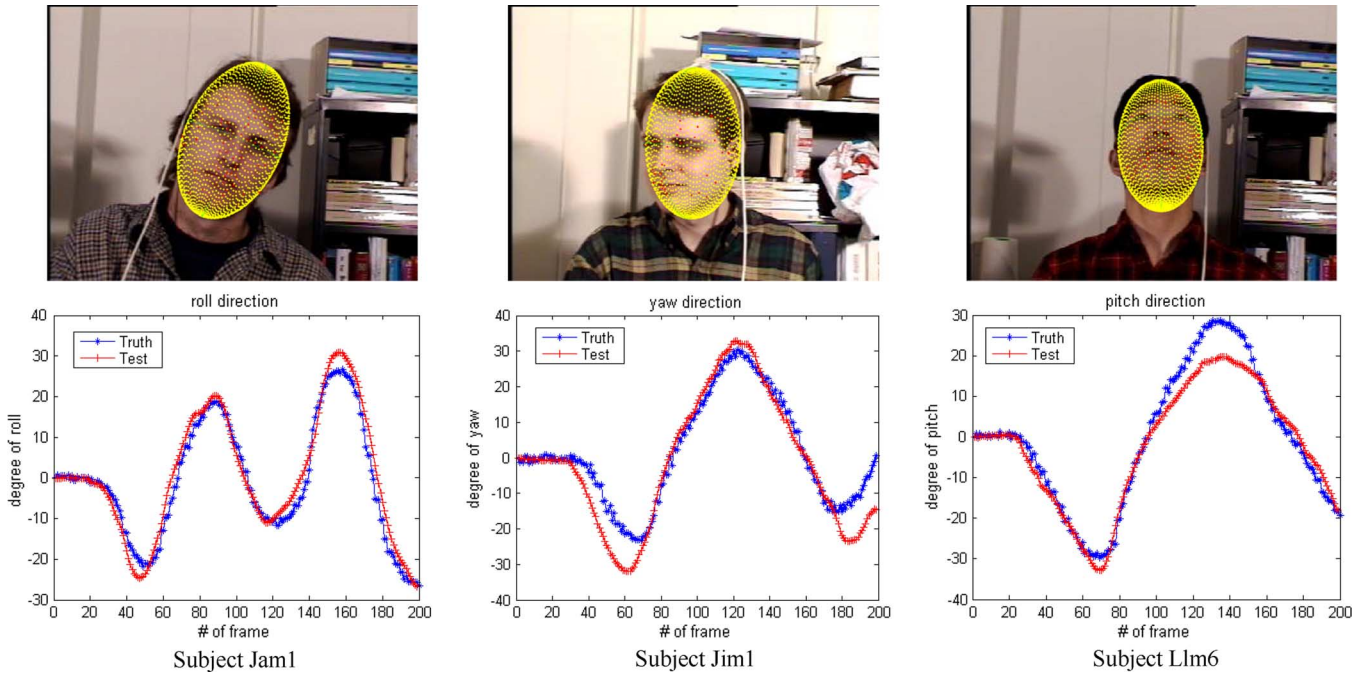


Fig. 6. Example of tracking result of roll, yaw, and pitch direction; the first row shows the fitting results and the second row displays the tracking results.

TABLE I
COMPARISON WITH PREVIOUS METHODS

	Proposed method	An's method [34]	Xiao's method [29]
Database	45 image seq. [27]	45 image seq. [27]	45 image seq.[27] + 5 image seq.[29]
Illumination	Uniform	Uniform	Uniform (45 seq.) & varying (5 seq.)
Model	3D ellipsoid	3D partial ellipsoidal	3D cylinder
Image size	320×240	320×240	320×240
Average error (°)	2.1°,3.6°,3.7°	2.8°,3.9°,4.0°	1.4°,3.2°,3.8°

fitting result and the second row displays the tracking result for an image sequence. In the first row, the yellow dots represent the model and the red dots are the selected SIFT features. In the second row, the blue line represents the ground truth and the red line shows the test results.

The average errors across the whole database in each direction are shown in Table I. The differences within the comparator results can be viewed as illusory in the sense that tracking results can depend on initialization. All methods show good performance (less than 4° error in each direction). However, the accuracy of the comparator results is perhaps compromised by the desire to achieve video-rate processing and these methods assume there is small variation of head pose between frames. Unlike other methods, our new method can be more robust under translation since our method is based on local feature matching first, and then analyzing the pattern of the face image. In our previous work [30], we presented another experiment at low frame rate (∼10 fps).

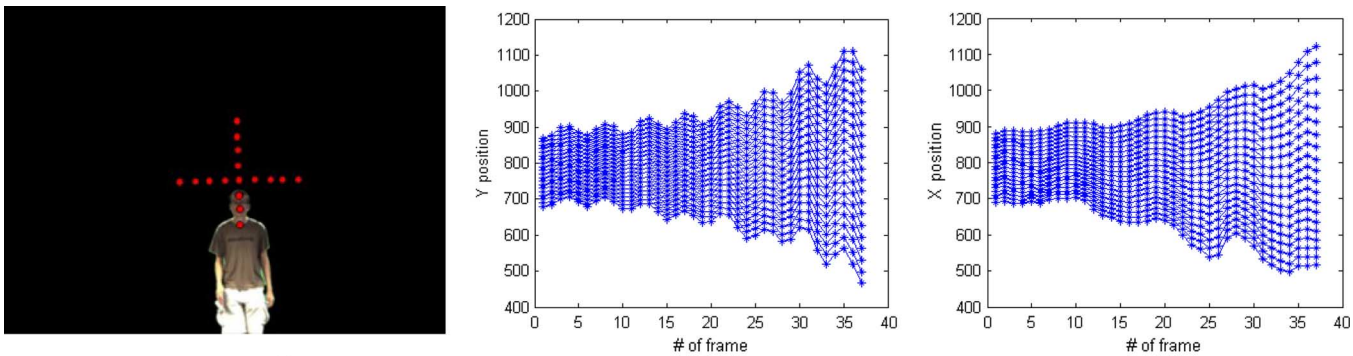
III. GAIT TRAJECTORY

From the previous section, the face pose can be extracted using the 3-D face model under a controlled environment particularly when the distance between the object and the camera is fixed. In reality, however, in visual surveillance environments, there are many changes in illumination and huge head movements. Also, the low frame rate could affect the detection rate. For example, the head movement might not be continuous and the illumination could change frame by frame, especially in a low frame rate video. Unfortunately, most CCTVs which are already installed in the public places have the above characteristics. To overcome the above difficulties we use not only a face image but also use alternative biometric information: gait.

A. Looming Field

The looming effect occurs when a subject walks toward a camera resulting in a nonlinear increase in apparent subject size [36]. Under the pinhole camera model [33], the relationship in perspective projection is shown in (11). In the relationship, when the walking position Z is changed linearly the projected position (x, y) is changed in nonlinear way. To show this we extracted the trajectory around the head in the following way. First, the background subtraction image can be extracted [37], and then the corresponding points between adjacent images are extracted. The same feature refinement methods are applied as Section II-A (the region- and distance-based methods). To determine the potential trajectory of the specific point we calculated the 2-D Homography relationship for each adjacent image. Then, as shown in Fig. 7(a), the x position is fixed for the first frame and then, the successive corresponding y axis trajectories are extracted by using the Homography matrices. Also, the y position is fixed and the trajectories of the successive corresponding x positions are found. Fig. 7 shows the trajectories of the above points. These are affected by the looming effect. During walking, from some position, here we call 'looming center', the trajectory can vary in nonlinear manner.

The trajectory from the looming center shows the constant variation (gradient is around zero) and the variation according



(a) (b) (c)

Fig. 7. Tracking result at the specific points; (a) chosen points, (b) the trace of Y position at the fixed X position, and (c) the trace of X position at the fixed Y position.

the frame number shows dependency proportional to the distance between the looming center and the chosen points. In the case of Fig. 7, we found that the looming center is located in around (780,800) in x and y axes respectively. Note that the looming center is not the center of the image (the image size is 1280 $\times$ 1024 pixels).

B. Gait Feature Extraction

To understand looming, we need to describe the gait trajectory. When a person is walking the movement of the head is conspicuous and sinusoidal [38]. Fig. 8 represents the trajectory of the head position for each frame. As a person walks towards the camera, the variation of the head movement in the vertical (y) direction increases. It also reveals that there is periodic movement in y direction. To clarify these facts we extract the exact human trajectory using some manually chosen points such as the points below the neck or the points in the chest for all image sequences. We use 35 samples (26 males and 9 females, with around 40 images in each sequence) chosen randomly from the Biometric tunnel database [24] and extract the corresponding points manually for all frames. Fig. 9 shows the results for the horizontal and vertical trajectories, respectively. Clearly, the vertical gait trajectory has a consistent trend. Its nature depends on the distance between the looming center and the position of the tracked pixel. Unlike the vertical trajectory the horizontal gait trajectory changes with a subject's gait. Also, the variation between points in the vertical trajectory is much larger than the variation in the horizontal trajectory. Therefore, we shall ignore the variation of the horizontal gait trajectory. In the next section we shall generate the model of the vertical gait trajectory and show the performance of model fitness using a nonlinear optimization method.

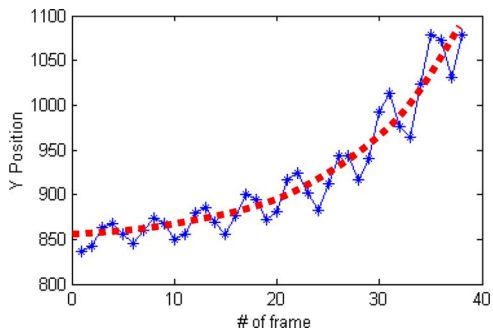


Fig. 8. Sample gait trajectory.

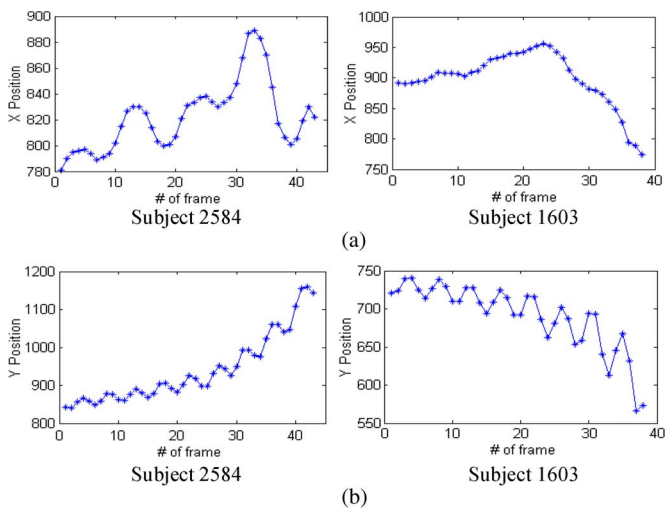


Fig. 9. Samples of the gait trajectory. (a) Horizontal variation of neck point; (b) vertical variation of neck point.

C. Gait Trajectory Model Definition

To define a gait trajectory model we use the following assumptions.

1) The variation in the z direction (walking direction) should be much larger than that in the x and y directions.
2) The walking speed and the sampling rate are constant.

In Section III-B the gait trajectory can be divided into two parts: a periodic factor and a scaling factor. Under the above constraints, the periodic factor increases or decreases in the same ratio. Therefore, the gait trajectory model could be defined as the following way, where the vertical position $y(t)$ is a function of gait frequency ω and $p(t), r(t)$ are the periodic factor and the scaling factor. First of all, the general case is

$$y(t) = p(t) \times \sin(\omega t + \theta) + r(t) \tag{10}$$

Under the pinhole camera model, the relationship in perspective projection [33] is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} fX/Z \\ fy/Z \end{pmatrix} \tag{11}$$

where (x, y) is a point in the 2-D image plane, (X, Y, Z) are the coordinates in 3-D space, and f is focal length. This is a conversion between 3-D camera coordinates and 2-D image coordinates.

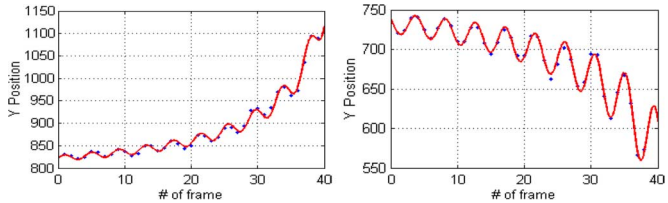


Fig. 10. Samples of the fitting result and autocorrelation of the error.

Assuming that the velocity of walking is a constant we can approximate the relationship between the walking direction and the time as

$$Z(t) = \mu - \delta t \quad \mu, \delta > 0 \quad Z(t), t \geq 0 \tag{12}$$

where $Z(t)$ represents the distance from the camera, μ is a track length, δ is a walking speed, and t is time or frame number. In 3-D space, the height of human is constant and the trajectory of the upper body shows the periodic function. Therefore, the vertical trajectory can be modeled in 3-D space.

$$Y(t) = C_1 \sin(\omega t + \theta) + C_2 \tag{13}$$

Substituting (12) and (13) into (11), we can obtain a simple relation as shown in (14).

$$y(t) = \frac{k_1}{\mu(1 - t/\lambda)} \sin(\omega t + \theta) + \frac{k_2}{\mu(1 - t/\lambda)} \tag{14}$$

where θ is the initial phase, λ is a total time, $\lambda = \mu/\delta$, and C_n, k_n are constants. Therefore, the scaling and the periodic factor can be defined using (14) which describes the general case of gait trajectory. It shows the sinusoidal wave which has nonlinear magnitude and nonlinear line equation; however, this equation does not handle the case that the tracked point is located around the looming center (in Fig. 7, the gradient around the looming field is about zero). Around the looming center, the scaling factor can be modeled as

$$scaling\ factor : \mathrm{g}_c(t) = C_o t + I_o \tag{15}$$

where C_o is around zero and I_o is the initial vertical position of the trajectory. Generally, average adult walking velocity on level surfaces is approximately 80 m/min. For men and women, it is about 82 m/min and 79 m/min, respectively [40]. While investigating the gait trajectories we can find that one period of gait trajectory contains four to five gait trajectory points. For example, in Fig. 8, there are four or five sampling points in a gait cycle. So, we set the constraint of gait frequency as from 1/5 to 1/4.

To evaluate the performance of the model we normalized all of the extracted gait trajectories. After fitting using the Levenberg-Marquadt algorithm, we use R-squared and Sum of Square Error (SSE). Fig. 10 shows the model fitting results for a typical gait trajectory and its error analysis. The blue points are the gait trajectory and the red line is the result of model fitting. Another advantage of this model is that it can express the trajectory between the trajectory points. The value of R-squared for all samples is 98.0% and the SSE is 3.66%. In addition, this model was used for gait application in our previous research [31], which showed good performance in the visual surveillance environment of the PETS 2006 data.

Fig. 11. Example of the approximated face region by gait trajectory and fitting results according to the frames; first sample is highly affected by lighting, second sample has large variation of yaw direction, last one has head pose variation.

D. Experimental Results

In initialization, first, the SIFT points are extracted from the background subtracted images. Each 2-D Homography matrix can be calculated from eight randomly selected points. The Direct Linear Transformation (DLT) [33] is used to solve the 2-D Homography from the corresponding points. To find the optimized Homography matrix RANdom SAmple Consensus (RANSAC) [39] is applied to choose the Homography matrix which has the largest number of inliers. Then, we selected a center point of face manually in the first frame in order to obtain a successive potential trajectory by the Homography matrices. After that, the gait trajectory model is applied to the potential trajectory. It results in generating a clear gait trajectory. Since the Homography between the frames is known, we can extract the approximate face region in each frame after initializing the height and width of the face region for the first frame. This initialization process can be automatic if the person looks towards the camera and the size of face image is sufficiently large to use a face detector; however, we assume the general case where these conditions cannot be guaranteed. In tracking, first the 3-D ellipsoidal model was applied in the first frame using a manually given motion vector. For the next frame, the motion vector is calculated using the valid features and nonlinear optimization.

By this procedure, we tested the 34 sequences from the Biometric tunnel database. Fig. 11 shows the experimental results. The first row per subject shows the result of extracting the approximated face regions by using the gait trajectory. The second

TABLE II
TEST RESULTS WITH/WITHOUT THE GAIT TRAJECTORY

	1st test	2nd test	3rd test
Image size (pixels)	1600×1200 (Raw image)	300×300 (Approximate face region)	1600×1200 (Raw image)
Algorithms	Viola & Jones	Viola & Jones	New method
Detection rate (%)	80.8% (957/1185)	94.0% (1114/1185)	98.0% (1161/1185)

row shows the tracking results which are displayed by the 3-D model fitting. As shown in Fig. 11, the background changes as the subject walks; however, the center and the size of the face are not changed. We can extract the face region regardless of pose variation, illumination change, and low resolution. The face region and pose is clearly extracted well, especially the comparison with the data from which it was extracted (Fig. 1).

To confirm the effectiveness of using the gait trajectory for face acquisition we test the face detection ratio with and without the gait trajectory model. Basically, the 3-D ellipsoidal model cannot be applied directly to the raw image due to the image resolution. To use the ellipsoidal model, a minimum image size should be guaranteed. So, in the above experiment we resize the approximate face region. An alternative way is to deploy the Viola-Jones face detection system [41]. We test three cases; *i*) face detection ratio with raw images using [41], *ii*) one with the approximate face regions using [41], and *iii*) one using only the gait trajectory. Table II shows the result of face detection.

As shown in Table II, using the gait trajectory shows a better face detection performance. Comparing all the tests (with and without the gait trajectory), it shows a 17.2% improvement beyond the use of Viola-Jones face detection. The second test shows detection aided by the gait trajectory. Ideally, the results of first and second test should show the same result. However, even though the resizing ratio is the same (default 1/1.2), the number of resized pixels differs. Therefore, this result also indicates the usefulness of the gait trajectory model. The differences between the second and the third test are caused by variation in head pose and illumination, confirming that the new method is indeed more robust than the conventional approaches in this environment.

IV. FRONTAL FACE IMAGE EXTRACTION

Face image extraction is the basic step for face recognition. In the previous Section, we extracted the approximate face regions and calculated the position of the face based on the 3-D ellipsoidal model and the gait trajectory model. Deployment depends on imaging scenario. For example, the Boston database was recorded under uniform illumination and at fixed distance. Even though the variation of head pose is large, the frontal face can be extracted since every image sequence contains the front pose of the face image. Thus, in this database, the head inclination is the main factor to consider when registering face image. In the case of the Biometric tunnel database where the person walks toward the camera, the image resolution of the face changes with time. Normally, when the person walks, the head rotates less than $\pm 15^{\circ}$ from the walking direction. Thus, the most important consideration is the resolution of the extracted face image rather than the face rotation. Considering all of possibilities, in this Section we will describe the three step preprocessing extraction methods leading to face recognition analysis.

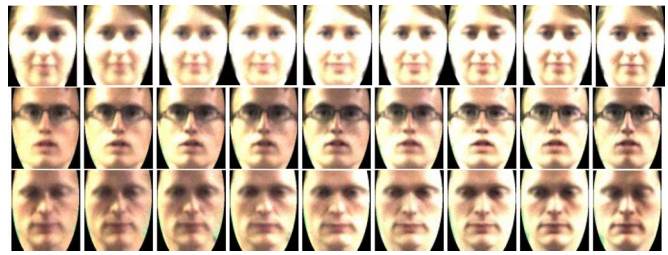


Fig. 12. Examples of reconstructed frontal face images.

A. Pose Based Face Image Filtering

There are many factors which are to be considered in face extraction such as the resolution of face image, facial expression, illumination, or facial obstacles in the face. In this research, we assume that there are no illumination changes, no obstacles, and no facial expressions. We are only concerned with the relationship between head pose, resolution, and face recognition. Ideally, the best sample for face recognition is a face image in which the face is looking directly towards the camera and the distance is the closest among all the images. In other words, the three axis angles $(\theta_{3Dx}, \theta_{3Dy}, \theta_{3Dz})$ are near zero and translation (t_{3DZ}) in the z axis direction is the smallest in (8).

B. Image Acquisition Using View Change

In previous section, we calculated the motion vector using the 3-D ellipsoidal model. Here, we extract and reconstruct the frontal face image with projecting the fitted 3-D ellipsoidal model into the 2-D image plane. Basically, if there is a large change in head pose, the reconstructed face image might be greatly distorted since our model is not based on an actual face model such as 3DMM. Also, we only consider the face images with pose within $\pm 10^{\circ}$, consistent with walking subjects, and where the frontal face can be reconstructed. Fig. 12 shows the results of the reconstructed frontal face images for images which have pose variation within $\pm 10^{\circ}$ from the first frame. The total number of images is around 40 images per subject.

C. High Resolution Image Reconstruction

To improve the resolution of face images we have deployed Super Resolution (SR) methods. In many surveillance images the face region occupies less than one tenth of the whole image so that the SR method can be used to improve the face pattern and texture for human interpretation. Also, SR can be deployed to extract more face features from the resized face image and to reconstruct an accurate face image from the synthesized frontal face images which are generated by the view change in the 3-D face models. The Low Resolution (LR) image sequences in visual surveillance environments could be used to reconstruct a High Resolution (HR) image.

To synthesize a HR image from LR images we then applied four of the existing SR methods: Bilateral Shift and Add (BSA) [44], Iterative Norm 1(IterNorm1) [42], Median Shift and Add (MSA) [44], Shift and Add (SA) [43]. The size of the LR images is around 70 × 90 to 100 × 130 pixels (the distance be-

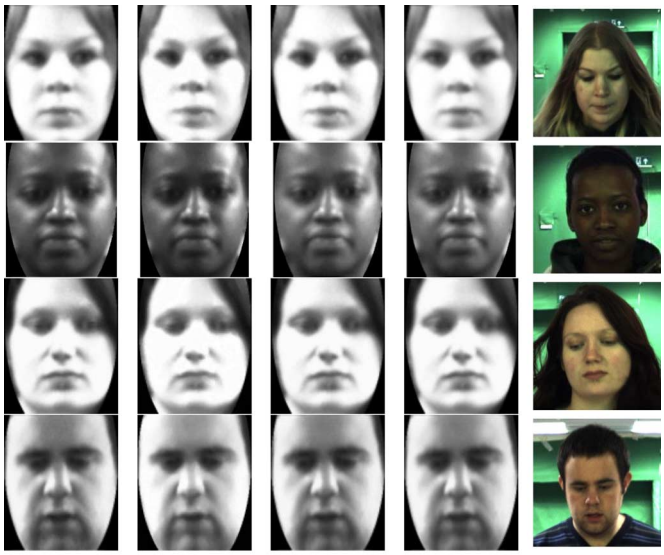


Fig. 13. Result of HR images and the last frame of the image sequences; HR images which are generated by (a) Bilateral Shift and Add method, (b) Iteration Norm 1 method, (c) Median Shift and Add method, and (d) Shift and Add method. (e) The last frame of the image sequence.

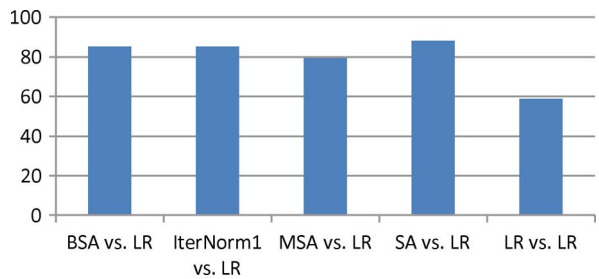


Fig. 14. Recognition rate between four HR images made by different SR methods and the other of LR images which were not used to build the HR images, using PCA-based recognition.

tween eyes' centers is around 30 to 40 pixels). To ease alignment we resized all LR image to be 100×130. The resulting HR images are two times larger in each direction than LR images. Fig. 13 shows the synthesized HR image for each method and the approximated face region in the last frame (the right-most column, Fig. 13(e)). The image, Fig. 13(e) is the highest resolution image acquired from the sequence as the subject is closest to the camera, and it was not used to derive the HR images. The last frame is shown for comparison between HR images and a real image. Images in which the subjects do not face the camera cannot be used for recognition. However, the pose-corrected HR image can be used for recognition. To evaluate recognition performance we analyze one sequence of around 40 images for each 34 subjects. We compare the HR image with the rest of frames from the sequence using Principal Component Analysis (PCA) based recognition [26], to give a baseline analysis of recognition potential. The gallery set consists of 502 pose-corrected images extracted from the same video, but these images are not used to generate the HR images. The probe set contains 136 HR images. Fig. 14 shows the resulting recognition rate between HR images and LR images. Of the super resolution techniques, generation by SA shows the highest recognition rate (88.2%) and generation by MSA shows the lowest recognition rate (79.4%). The average recognition rate is 84.6%. This indicates that the HR images can be successfully generated from the LR images with fidelity that is sufficient for recognition purposes. In order to confirm that using our approach is more effective for face recognition than using the LR images, we test the recognition rate between 340 LR images which are used to generate the HR images and the remaining 502 LR images which are not used to generate the HR images. The column furthest on the right in Fig. 14 shows that the resulting recognition rate is 58.8%, which is around 30% lower than that for HR images, confirming the advantages of this new approach.

The average recognition rate is over 84% for the HR images. Considering constraints such as large changes in illumination and low resolution, the recognition rate is high. Besides that, there are no directly comparable results generated by other techniques possible since this is the first approach to apply SR techniques using a 3-D face model and gait from image sequences in a surveillance environment.

V. Conclusions

We describe a new method of frontal face extraction by 3-D head pose estimation, gait trajectory analysis, and super-resolution analysis. We have shown how the head movement can be modeled by gait to allow for accurate face extraction when a subject walks towards a camera. Head pose estimation combines a 3-D ellipsoidal model with SIFT-based low-level feature extraction; the gait model analyses the trajectory of a looming subject. The approaches have been demonstrated with a Biometric tunnel database. The model fitting with actual data showed that the matching was over 98%. Also, SR methods of monochrome image sequences were applied to the reconstructed frontal face images. PCA based recognition was used as a measurement to evaluate the quality of HR images. By the new method the overall recognition performance is 84.6% whereas it was around 58.8% for the images from which the super-resolution images were derived. From the proposed methods, face images suitable for recognition can now be derived in visual surveillance environments. In the future we aim to translate these approaches to analyze surveillance data to provide a high resolution face image from video data in which a subject's movement is unconstrained.

References

[1] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 831–846, May 2010.

[2] G. Medioni, J. Choi, C. H. Kuo, and D. Fidaleo, "Identifying noncooperative subjects at a distance using face images and inferred three-dimensional face models," *IEEE Trans. Syst. Man, Cybern. A, Syst. Humans*, vol. 39, no. 1, pp. 12–24, Jan. 2009.

[3] F. W. Wheeler, X. Liu, and P. H. Tu, "Multi-frame super-resolution for face recognition," in *Proc. IEEE Conf. Biometrics, Theory, Appl., Syst.*, 2007, pp. 1–6.

[4] P. Mortazavian, J. Kittler, and W. Christmas, "A 3-D assisted generative model for facial texture super-resolution," in *Proc. IEEE Conf. Biometrics, Theory, Appl., Syst.*, 2009, pp. 452–458.

[5] K. Jia and S. Gong, "Generalized face super-resolution," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 873–886, Jun. 2008.

[6] P. H. H. Yeomans, S. Baker, and B. V. K. V. Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, 8 pp.

[7] W. W. W. Zou and P. C. Yuen, "Learning the relationship between high and low resolution images in kernel space for face super resolution," in *Proc. Int. Conf. Pattern Recog.*, 2010, pp. 1152–1155.
[8] B. L. H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 20–23, Jan. 2010.
[9] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vis. Image Underst.*, vol. 101, no. 1, pp. 1–15, Jan. 2006.
[10] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, "3D assisted face recognition: A survey of 3D imaging, modeling and recognition approaches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, 11 pp..
[11] U. Park and A. K. Jain, "3D model-based face recognition in video," in *Proc. IEEE/IAPR Int. Conf. Biometrics*, 2007, pp. 1085–1094.
[12] S. V. Duhn, M. J. Ko, L. Yin, T. Hung, and X. Wei, "Three-view surveillance video based face modeling for recognition," in *Proc. Biometric Symp.*, 2007, 6 pp.
[13] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.
[14] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
[15] C. T. Liao, S. F. Wang, Y. J. Lu, and S. H. Lai, "Video-based face recognition based on view synthesis from 3D face model reconstructed from a single image," in *Proc. IEEE Int. Conf. Multimedia, Expo*, 2008, pp. 1589–1592.
[16] R. Stiefelhagen, K. Bernardin, H. K. Ekenel, and M. Voit, "Tracking identities and attention in smart environments—Contributions and progress in the CHIL project," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2008, 8 pp.
[17] S. Argyropoulos, D. Tzovaras, D. Ioannidis, and M. G. Strintzis, "A channel coding approach for human authentication from gait sequences," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 3, pp. 428–440, Sep. 2009.
[18] J. Lu and Y. Tan, "Gait-based human age estimation," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 761–770, Dec. 2010.
[19] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, "The effect of time on gait recognition performance," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 543–552, Apr. 2012.
[20] A. Kale, A. K. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *Proc. IEEE Int. Conf. Acoustics Speech, Signal Process.*, 2004, pp. 901–904.
[21] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: Data sets, performance and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
[22] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 439–446.
[23] X. Zhou and B. Bhanu, "Integrating face and gait for human recognition at a distance in video," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1119–1137, Oct. 2007.
[24] R. D. Seely, S. Samangooei, L. Middleton, J. N. Carter, and M. S. Nixon, "The university of southampton multi-biometric tunnel and introducing a novel 3D gait dataset," in *Proc. IEEE Conf. Biometrics, Theory, Appl., Syst.*, 2008, pp. 1–6.
[25] I. Matthews and S. Baker, "Active appearance models revisited," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 135–164, Feb. 2004.
[26] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1991, pp. 586–591.
[27] M. La Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 322–336, Apr. 2000.
[28] S. Basu, I. Essa, and A. Pentland, "Motion regularization for model-based head tracking," in *Proc. Int. Conf. Pattern Recog.*, 1996, pp. 611–616.
[29] J. Xiao, T. Kanade, and J. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2002, pp. 156–162.
[30] S. U. Jung and M. S. Nixon, "On using gait biometrics to enhance face pose estimation," in *Proc. IEEE Conf. Biometrics, Theory, Appl., Syst.*, 2010, pp. 1–6.
[31] S. U. Jung and M. S. Nixon, "Detection human motion with heel strikes for surveillance analysis," in *Proc. Int. Conf. Comput. Anal. Image. Patterns*, 2011, pp. 9–16.
[32] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. Boca Raton, FL: CRC Press, 1994.
[33] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[34] K. H. An and M. J. Chung, "3D head tracking and pose-robust 2D texture map-based face recognition using a simple ellipsoid model," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots, Syst.*, 2008, pp. 307–312.
[35] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Jan. 2004.
[36] T. K. M. Lee, M. Belkhatir, and S. Sanei, "Fronto-normal gait incorporating accurate practical looming compensation," in *Proc. Int. Conf. Pattern Recog.*, 2008, pp. 1–4.
[37] G. Cheung, T. Kanade, J. Bouquet, and M. Holler, "A real time system for robust 3d voxel reconstruction of human motions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2000, pp. 714–720.
[38] Aristotle, On the Motion of Animals, B.C. 350 [Online]. Available: http://classics.mit.edu/Aristotle/motion_animals.html 15/4/2004
[39] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[40] J. M. Burnfield and C. M. Powers, *Normal and Pathologic Gait, in Orthopaedic Physical Therapy Secrets*, 2nd ed. St. Louis, MO: Elsevier, 2006.
[41] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, Feb. 2001.
[42] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
[43] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.
[44] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," in *Proc. Appli. Digital Signal, Image Process.*, 2003, pp. 121–130.

Sung-Uk Jung received the B.Sc. degree in electrical engineering from Korea University in 2003, the M.Sc. degree in electrical engineering and computer science from Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2005, and the Ph.D. degree in electronics and computer science from the University of Southampton, U.K., in 2012.

Since August 2005, he has been with Electronics and Telecommunications Research Institute (ETRI), Republic of Korea, and is currently with the Human Identification Research Team of ETRI. His current research interests include embedded computer vision, biometric systems, human motion analysis, intelligent video surveillance, and intelligent robot.

Mark S. Nixon is a Professor in computer vision at the University of Southampton, Southampton, U.K. His research interests are in image processing and computer vision. His team develops new techniques for static and moving shape extraction which have found application in automatic face and automatic gait recognition and in medical image analysis. His team contains early workers in face recognition, and they later came to pioneer gait recognition and more recently joined the pioneers of ear biometrics. His vision textbook, with A. Aguado, *Feature Extraction and Image Processing* (Academic) reached second edition in 2008. With T. Tan and R. Chellappa, the book *Human ID Based on Gait* is part of the Springer Series on Biometrics and was published in 2005. He has chaired or had major involvement in many conferences (BMVC, AVBPA, IEEE Face and Gesture, ICPR, ICB, IEEE BTAS) and has given many invited talks.

Dr. Nixon is a Fellow IET and FIAPR.